Citation: López, R., Barragán, J., Colomer, M.A., 2007. Reply to the comment by Rupp and Smart on ‘‘Flow resistance equations without explicit estimation of the resistance coefficient for coarse-grained rivers,’’ *J. Hydrol.*, 346 (3–4), 179–183. https://doi.org/10.1016/j.jhydrol.2007.08.023.

Reply to the comment by D.E. Rupp and G.M. Smart on "Flow resistance equations without explicit estimation of the resistance coefficient for coarse-grained rivers"

Raúl López

Department of Agroforestry Engineering, University of Lleida, Av. Alcalde Rovira Roure, 191, E25198, Lleida, Spain. Tel. +34 973 70 28 20. Fax +34 973 70 26 73. E-mail: rlopez@eagrof.udl.es. Corresponding author.

Javier Barragán

Department of Agroforestry Engineering, University of Lleida, Av. Alcalde Rovira Roure, 191, E25198, Lleida, Spain. E-mail: barragan@eagrof.udl.es.

M. Àngels Colomer

Department of Mathematical, University of Lleida. C/ Jaume II, 69 Campus Cappont, E25001, Lleida, Spain. E-mail: colomer@matematica.udl.es

Introduction

Rupp and Smart comment on a paper by López et al. (2007) in which a series of models were calibrated and validated (with data from coarse-grained rivers), with the aim of providing equations for which no explicit estimate is needed for the term of flow resistance. This objective has been tackled with similar methods in previous studies (among the most recent, Dingman and Sharma, 1997; Bjerklie et al., 2003; Bjerklie et al., 2005). The general form of the models cited is

$$Q = KA^{\delta}R^{\alpha}S^{\beta} \tag{1}$$

where Q is discharge, K is a coefficient, A is the cross-sectional area, R is the hydraulic radius, S is the friction slope, δ and α are exponents, and β is either an exponent or a function.

In Rupp and Smart's opinion, two principal criticisms can be made of the above focus. These are, in first place, the little attention paid to the theoretical basis of Eq. (1) and, secondly, the effect that the severe multicollinearity has on the parameters fitted through multiple regression. Based on these, Rupp and Smart propose an alternative model, which, in their opinion, has a greater physical basis and avoids the prejudicial effects of multicollinearity. The aim of this reply is to respond to the questions posed and to fit and compare alternative models (considered in the light of Rupp and Smart's comments) to those calibrated in López et al. (2007).

Theoretical considerations

One of the main theoretical questions that Rupp and Smart pose in their comment is the explanation of the value of the fitted parameters of Eq. (1) through the adoption of a specific law of vertical distribution of flow velocity. The integration of this law on the flow depth allows a function of the mean flow velocity (V) to be obtained.

Laws of velocity distribution, such as the logarithmic or the one proposed by Aguirre-Pe and Fuentes (1990), are applicable to a wide range of relative submergence, but may not be valid under very high relative roughness conditions (e.g., when not all the roughness elements are submerged). Under large-scale roughness conditions, different shapes of velocity profile have been observed: quasi-logarithmic, linear, irregular and S-shaped (Byrd et al., 2000); and inflectional (e.g., Katul et al., 2002) and S-shaped (e.g., Ferro and Pecoraro, 2000) models have been proposed, among others. Exponential, linear and uniform models have been suggested specifically for the near-bed region (between roughness troughs and roughness tops) (Nikora et al., 2004). However, it is not feasible to obtain a law of velocity distribution with a simple equation, and one which is also valid for the full range of common relative submergence in gravel-bed and mountain rivers and for all flow specific layers.

Despite this, the full logarithmic law of resistance to flow (Smart et al., 2002) is, in practice, possibly one of the most recommendable to cover a wider range of flow relative submergence. This equation forms part of the theoretical basis of Eqs. (9) and (10), proposed as alternatives by Rupp and Smart in their comment. However, if the characteristics of the fitted database fall significantly outwith the theoretical restrictions of the full log law, in practice the equation obtained can be considered more as an empirical approach than a physically-based model. This occurs to a significant degree with the database selected for the study by López et al. (2007). For example, although the selection of data in the above-mentioned paper was restricted to non-sinuous, quasi-prismatic and non-vegetated channel reaches, 57% of this data had a y/d_{84} value (where y is the mean flow depth and d_{84} is the grain size for which 84% of bed material is finer) lower than 4, 39% below 2 and 16% lower than 1. In such relative submergence conditions, which are frequent in mountain rivers, the validity of the full logarithmic law may be questionable. In this sense, the data set taken in New Zealand rivers, to which Eqs. (9) and (10) are fitted in the comment by Rupp and Smart, also incorporates the effects of bank vegetation, sinuosity and the longitudinal variation of the channel cross-section (Hicks and Mason, 1991; Smart, 2004).

Moreover, given that the main aim is to have equations with no explicit resistance term, in Rupp and Smart's comment, the relationship $z_0 = K_1 S^{\gamma}$ (where z_0 is the hydraulic roughness of the boundary and K_1 and γ are fitted parameters) is adopted in Eq. (10) (given that no evidence is found for the relation of z_0 with R or A). As the mentioned researchers state, this relationship lacks a physical basis, which in fact means introducing a higher degree of empirical approach. In short, if it is borne in mind that an important proportion of the fitted database breaks the theoretical restrictions of the full log law and that, moreover, z_0 is related to S empirically, the final result could be considered closer to an empirical relationship than to a physically-based model.

Another of the theoretical questions that Rupp and Smart deal with in their comment refers to the theoretical value of the exponent of S. The Chézy equation can be derived from the force-balance relations and accepting that the shear stress on a channel boundary is proportional to the square of the flow velocity (V^2). Thus, the exponent of S is equal to 0.5. However, Leopold et al. (1960) point out that the resistance can only be proportional to V^2 if the flow boundary does not change substantially with discharge. This requirement is generally met for pipe flow but not for open channel flow (Bjerklie et al., 2005). Leopold et al. (1960) hypothesized that internal distortion resistance and spill resistance produce deviations from the velocity-squared law. Based on the analysis of the flume data presented in Leopold et al. (1960) and previous empirical studies, Bjerklie et al. (2005) suggested a value of the exponent of S equal to 0.33. For a river database, they found that if a constant value of K (Eq. 1) was adopted, the Chézy and Manning equations modified by imposing $\beta = 0.33$ reached a better fit than the original equations (i.e., $\beta = 0.5$). Moreover, they confirmed that the predictive power of the modified Manning equation and the one obtained by fitting the coefficient and exponents of Eq. (1) are similar.

In this context it is worth indicating that in Eq. (10), proposed by Rupp and Smart, S does not really intervene as a potential term with an exponent equal to 0.5, as by adopting the relationship $z_0 = K_1 S^{\gamma}$ the influence of S is more complex. The intervention of R is also more complex, as it appears in various terms, in addition to the potential term with an exponent equal to 0.5.

Multiple regression considerations

Given that the multicollinearity (correlation between predictor variables) is a problem related to the sample, there are no contrast methods strictly speaking, but rather different practical procedures are used that estimate the level of multicollinearity (given that the most common situation is that there is a certain level of correlation between the predictor variables). One of these procedures consists of analyzing the value of the simple correlation coefficients between the explanatory variables two by two, with high values of the coefficient indicating a high multicollinearity. For the data set in López et al. (2007), the simple correlation coefficient (r) between log$A$ and log$R$ is 0.94, between log$A$ and log$S$ it is –0.57 and between log$R$ and log$S$ it is –0.50. However, in the field of multiple regression models, the above procedure is not the most recommendable, given that the correlation coefficients only include the correlation between pairs of explanatory variables.

To avoid the above, the Variance Inflation Factor (VIF) is frequently used. This factor measures directly how much the variance of each coefficient is inflated as compared to a situation with uncorrelated independent variables. The VIF is computed as: $\mathrm{VIF}_j = 1/(1 - R_j^2)$, where $R^2{}_j$ is the coefficient of determination obtained from carrying out the linear regression of the j-th predictor variable on the rest of the predictor variables in the model. The VIF values obtained with the database used in López et al. (2007) and for $\log A$, $\log R$ and $\log S$ are 9.26, 8.33 and 1.50 respectively. There is no threshold value for the VIF above which it can be affirmed that there are serious problems of multicollinearity. Some statisticians frequently consider that there is a serious multicollinearity problem when the value of the VIF is higher than 10 (Kleinbaum et al., 1988), although others state that the consequences can already be relevant for values above 5. None of the VIFs calculated surpasses the threshold of 10 (although those corresponding to $\log A$ and $\log R$ are higher than 5) and the one corresponding to $\log S$ is much lower than 5. Similarly, Dingman and Sharma (1997) obtained VIF values of 6.77, 6.74 and 1.32, respectively for the three variables cited above.

According to another practical rule, significant changes in the value of the fitted parameters when small changes are made to the database (e.g., adding or removing a limited number of observations) indicate important multicollinearity effects. In this sense, it must be borne in mind that during the cross validation phase in López et al. (2007), two equations were obtained for each of the models represented by Eqs. (21) and (22) (the two models most susceptible to severe multicollinearity, owing to the adoption of δ, the exponent of A, as a fitted parameter). Bear in mind that the fitted subset used in the validation of each equation corresponds to the random splitting of the total database into equal parts, which evidently is more than a small change to this database. Despite this, it was found in both models that there are no significant differences at the 95% confidence level between the parameters of the two equations obtained in the cross validation, nor between these and the parameters of the equation fitted to the full database. This would indicate that the level of multicollinearity has no serious effects.

In short, and in function of the above, no firm opinion can be expressed about the seriousness of the multicollinearity corresponding to $\log A$ when a value fitted to δ is adopted, while it can be affirmed that the level of collinearity between $\log R$ and $\log S$ (whose VIF does not exceed 1.33) is low.

Comparison of alternative models

In this section, Eqs. (33) and (35) (López et al., 2007) are compared with those obtained by fitting alternative models (considered in light of the comment by Rupp and Smart) to the database ($N = 904$) compiled in López et al. (2007). The alternative models analyzed in this section are those numbered as Eqs. (9)–(11) in the comment by Rupp and Smart and the modified Manning equation, proposed in Bjerklie et al. (2005), these being

$$Q = \frac{\sqrt{g}}{\kappa} A R^{1/2} S^{1/2} \left\{ \left[\frac{1}{1-(z_0/R)} \right] \ln\left(\frac{R}{z_0}\right) - 1 \right\} \quad (2)$$

$$Q = \frac{\sqrt{g}}{\kappa} A R^{1/2} S^{1/2} \left\{ \left[\frac{1}{1-(K_1 S^{\gamma}/R)} \right] \ln\left(\frac{R}{K_1 S^{\gamma}}\right) - 1 \right\} \quad (3)$$

$$Q = K_0 A R^{\alpha} S^{1/2} \quad (4)$$

$$Q = K_0 A R^{2/3} S^{1/3} \quad (5)$$

The parameters of the logarithmic transformation of Eqs. (2), (4) and (5) were estimated by a non-linear optimization algorithm that minimizes the sum of the squared residuals, where Q (in m^3/s), A (in m^2), R (in m) and S (m/m) are known and imposing $\kappa = 0.40$ and $g = 9.81$ m/s^2. In this case, it was not possible to apply the same procedure to the fit of Eq. (3). That was due to the fact that the fitted database includes an appreciable percentage of data in low relative submergence conditions. In fact, for the data with high relative roughness, and in function of the values of K_1 and γ considered to begin the optimization algorithm, the $R/(K_1S^\gamma)$ relationship can adopt values that are even lower than 1. This leads to predictions of negative values for Q and, thus, impedes the fit of logarithmic transformation of Eq. (3). An attempt has been made to solve this problem by expressing Eq. (3) in function of the dimensionless mean velocity (V/v_*)

$$\frac{Q/A}{\sqrt{gRS}} = \frac{V}{v_*} = \frac{1}{\kappa}\left\{\left[\frac{1}{1-(K_1S^\gamma/R)}\right]\ln\left(\frac{R}{K_1S^\gamma}\right) - 1\right\} \tag{6}$$

Eq. (6) was fitted through a non-linear optimization algorithm that minimizes the sum of the squared residuals, imposing $\kappa = 0.40$ y $g = 9.81$ m/s^2. It must be taken into account that certain degree of spurious correlation may occur, given that S appears on both sides of the equality of Eq. (6).

Table 1 shows the value of the fitted parameters for Eqs. (2)–(5) and Table 2 specifies the value of various fitted statistics (defined in López et al., 2007) corresponding to the above equations. For comparison with Eqs. (2)–(5), Eqs. (33) and (35), referred to in López et al. (2007), will be considered. Eq. (34) from the above-mentioned paper is not included in this analysis as it has no advantages in comparison with Eq. (33), given that the latter has a lower number of fitted parameters, its explanatory power is slightly higher and it is less susceptible to severe multicollinearity.

Table 1. Value of the fitting parameters

Parameter	Eq. (2)	Eq. (3)	Eq. (4)	Eq. (5)
K_0 ($m^{1\text{-}\alpha}s^{-1}$)	7.83[a]	7.83[a]	23.6	7.94
α	1/2[a]	1/2[a]	1.01	2/3[a]
β	1/2[a]	1/2[a]	1/2[a]	1/3[a]
z_0 (m)	0.0176	–	–	–
K_1 (m)	–	0.27[b]	–	–
γ	–	0.53[b]	–	–

[a] Fixed value.

[b] Parameters fitted to Eq. (6).

Table 2. Calibration statistics of fitted equations of alternative models

Statistic	Eq. (2)	Eq. (3)	Eq. (4)	Eq. (5)
N_p	1	2	2	1
SE (m^3/s)	153	179	307	129
R^2	0.899	0.908	0.876	0.922
E'	0.819	0.852	0.774	0.817
MRE (%)	53.0	34.3	45.7	43.3
RE_{25} (%)	45.6	52.0	44.8	43.6
RE_{50} (%)	73.8	79.2	75.6	80.5
MSE (%)	30.0	13.9	16.6	14.8
OE (%)	52	55	46	45
AIC	9,099	9,382	10,361	8,796
BIC	9,102	9,389	10,368	8,799
m (–)	–0.35	–0.49	–0.33	–0.62
R^2 (t.l.)	0.11	0.30	0.10	0.41

N_p = number of fitting parameters in the model.

m, R^2 (t.l.) = slope and coefficient of determination of the trend line of residuals of log-transformed equations against Froude number.

When contrasting the explanatory power and prediction error of the fitted equations, it must be borne in mind that, being based on the square of the differences between the observed and predicted values, the SE, R^2, AIC and BIC indices are more sensitive to the extreme values of the fitting set. The E' index diminishes this sensitivity on being based on the absolute value of residuals, but does not treat all the data equally independently of magnitude. For this reason, when the goodness-of-fit to the entire data set is analysed globally, those models that follow the observations of greater magnitude correctly may reach an artificially high value of the indices based on absolute residuals. This could obscure the true explanatory power of the model for the majority of the rest of the sample (Legates and McCabe, 1999). To avoid this, it is recommendable, in the field of multiple regression, to analyze the goodness-of-fit in various classes of the dependent variable. This analysis is not tackled in this work, although, as well as the above-mentioned indices, other types of statistics (MRE, MSE, RE_{25}, RE_{50} and OE) are included that assign the same weight to all data independently of its magnitude. This will help to give a more general view of the behaviour of the models over the full range of the dependent variable. For this current work, the statistics of the first type dealt with (SE, R^2, E', AIC y BIC) are called indices based on the absolute error and those of the second type (MRE, MSE, RE_{25}, RE_{50} and OE) are cited as indices based on the relative error.

If the fit achieved for Eqs. (2), (4) and (5) (Table 2) is compared with that reached by Eqs. (33) and (35) (Table 3 in López et al., 2007), it is seen that the first group of equations generally reaches a slightly, or moderately, better results of the indices based on absolute error. In the case of the AIC and BIC indices, the improvement is partly due to the fact that Eqs. (2), (4) and (5) have a lower number of fitted parameters. However, if the statistics related to the relative error are analysed, Eqs. (2), (4) and (5) obtain rather poorer results. For example, the MRE of Eqs. (2), (4) and (5) varies between 43% and 53% (between 31% and 60% higher than in the case of Eq. (35)) and the MSE oscillates between 15% and 30% (between 63% and 230% higher than in the case of Eq. (35)).

In contrast with Eqs. (2), (4) and (5), Eq. (3) shows a more similar fit for the two groups of statistics in comparison with Eqs. (33) and (35). A trend towards an improvement of the indices based on absolute error is seen, while the values of the statistics based on relative error are only slightly lower (although to a greater degree for MSE).

Table 2 shows the value of the slope and the coefficient of determination of the trend line for the data in the plane Froude number–residuals of log-transformed discharge for Eqs. (2)–(5). The values corresponding to Eq. (33) are –0.65 and 0.56, and –0.60 and 0.53 for Eq. (35). It can be seen that the highest values of the slope of the trend line and its coefficient of determination are found in Eqs. (33), (35) and in Eq. (5). The lowest values are given by Eqs. (2) and (4), whose exponent of S is equal to 0.5. However, note that Eq. (3) has significantly higher values than those in Eq. (4), closer to those in Eqs. (33), (35) and Eq. (5). This could be attributed to the fact that in reality S in Eq. (3) does not intervene only as a potential term with an exponent equal to 0.5.

In conclusion, among the equations studied (Table 1), Eq. (3), based on the logarithmic model of vertical velocity distribution, is the only one that reaches an explanatory power and goodness-of-fit comparable with those obtained by Eqs. (33) and (35) in López et al. (2007). As Eq. (3) has a smaller number of fitted parameters, this means a comparative advantage for the above-mentioned equation. However, the limitations and observations presented below must be borne in mind.

The fit of Eq. (3) to data sets that contain observations on very high relative roughness conditions ($R/d_i < 1$, where d_i is the grain size for which i% of bed material is finer) may be nonviable through the logarithmic transformation of the mentioned equation. It must also be taken into account that, strictly speaking, the physical basis of Eq. (3) requires some restrictions to the calibration set that are not met to an appreciable degree in the context of high-gradient streams. Thus, in function of the characteristics of the database, the fitted equation could be considered the product of a statistical approach, rather than physically based. Lastly, it must be taken into account that the variables R and S in Eq. (3) do not, in reality, intervene as potential terms with exponent 1/2, given these variables also intervene in a more complex way in other terms of the equation.

References

Aguirre-Pe, J., Fuentes, R., 1990. Resistance to flow in steep rough streams. J. Hydraul. Eng. 116 (11), 1374–1386.

Bjerklie, D.M., Dingman, S.L., Vorosmarty, C.J., Bolster, C.H., Congalton, R.G., 2003. Evaluating the potential for measuring river discharge from space. J. Hydrol. 278 (1–4), 17–38.

Bjerklie, D.M., Dingman, S.L., Bolster, C.H., 2005. Comparison of constitutive flow resistance equations based on the Manning and Chezy equations applied to natural rivers. Water Resour. Res. 41, W11502.

Byrd, T.C., Furbish, D.J., Warburton, J., 2000. Estimating depth-averaged velocities in rough channels. Earth Surf. Process. Landforms 25, 167–173.

Dingman, S.L., Sharma, K.P., 1997. Statistical development and validation of discharge equations for natural channels. J. Hydrol. 199 (1–2), 13–35.

Ferro, V., Pecoraro, R., 2000. Incomplete self-similarity and flow velocity in gravel bed channels. Water Resour. Res. 36 (9), 2761–2769.

Hicks, D.M., Mason, P.D., 1991. Roughness characteristics of New Zealand rivers. New Zealand Water Resources Survey, DSIR Marine and Freshwater, Wellington, NZ.

Katul, G., Wiberg, P., Albertson, J., Hornberger, G., 2002. A mixing layer theory for flow resistance in shallow streams. Water Resour. Res. 38 (11), 1250.

Kleinbaum, D.G., Kupper, L.L., Muller, K.E. 1988. Applied regression analysis and other multivariable methods. PWS-Kent Publishing Company, Boston, Mass., USA.

Legates, D.R., McCabe, G.J., 1999. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resour. Res. 35 (1), 233–241.

Leopold, L.B., Bagnold, R.A., Wolman, M.G., Brush, L.M., 1960. Flow resistance in sinuous or irregular channels. Geol. Surv. Prof. Pap. 282-D, US Geological Survey, Washington D.C., USA.

López, R., Barragán, J., Colomer, M.A., 2007. Flow resistance equations without explicit estimation of the resistance coefficient for coarse-grained rivers. J. Hydrol. 338 (1–2), 113–121.

Nikora, V., Katinka, K., McEwan, I., McLean, S., Dittrich, A., 2004. Velocity distribution in the roughness layer of rough-bed flows. J. Hydraul. Eng. 130 (10), 1036–1042.

Smart, G.M., Duncan, M.J., Walsh, J.M., 2002. Relatively rough flow resistance equations. J. Hydraul. Eng. 128 (6), 568–578.

Smart, G.M., 2004. An improved flow resistance formula. River Flow 2004–Proceedings of the Second International Conference on Fluvial Hydraulics. M. Greco, A. Carravetta and R.D. Morte (eds.), vol. 1, pp. 259-263, Balkema.